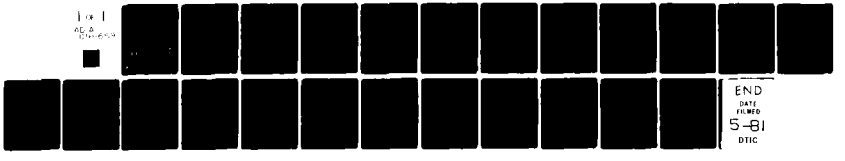
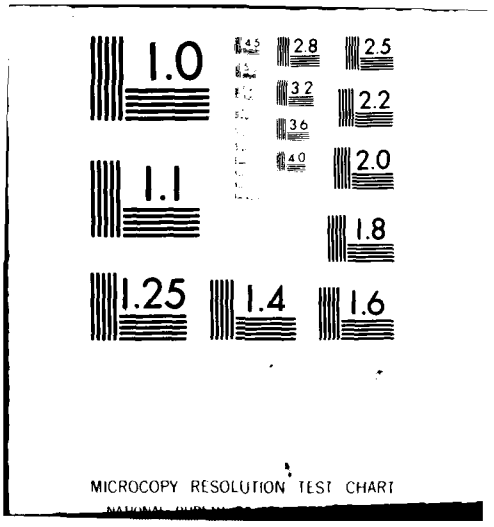
AD-A098 659 ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND ABERD--ETC F/G 9/2
A SURVEY OF LINEAR REGRESSION COMPUTER PACKAGES.(U)
FEB 81 R T MARUYAMA
UNCLASSIFIED ARBRL-TR-02289 SBIE-AD-E430 595 NL



END
DATE FILMED
5-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
MICROCOPY RESOLUTION TEST CHART

AD

TECHNICAL REPORT ARBRL-TR-02289

A SURVEY OF LINEAR REGRESSION COMPUTER PACKAGES

Richard T. Maruyama

February 1981

US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

Destroy this report when it is no longer needed.
Do not return it to the originator.

Secondary distribution of this report by originating or sponsoring activity is prohibited.

Additional copies of this report may be obtained from the National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia 22161.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

The use of trade names or manufacturers' names in this report does not constitute indorsement of any commercial product.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TECHNICAL REPORT ARBRL-TR-02289 ✓	2. GOVT ACCESSION NO. AD-A098659
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) A SURVEY OF LINEAR REGRESSION COMPUTER PACKAGES	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) RICHARD T. MARUYAMA	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Ballistic Research Laboratory ATTN: DRDAR-BLB Aberdeen Proving Ground, MD 21005	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS RDT&E 1L161102AH43
11. CONTROLLING OFFICE NAME AND ADDRESS USAARRADCOM Ballistic Research Laboratory ATTN: DRDAR-BL Aberdeen Proving Ground, MD 21005	12. REPORT DATE FEBRUARY 1981
	13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Statistical Computer Program
Stepwise Linear Regression
Curve Fitting
Linear Models

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (eal)

This report is a survey of a number of linear regression computer packages. Linear regression statistical methods are continually being utilized as a tool to analyze large sets of data. Because of this widespread utilization, a survey to research the range of linear regression packages now available is appropriate. A number of major statistical packages were selected for this survey and a comparison table of their statistical characteristics is included in this report. The outcome of this survey is that today's

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

major packages offer many additional options; some of the options, sometimes other subroutines, add a level of flexibility to the general regression analysis.

TABLE OF CONTENTS

I. INTRODUCTION

In August 1979, Dr. Norman R. Draper of the Mathematics Research Center, University of Wisconsin, conducted a three-day course titled "Regression Theory" sponsored by the U.S. Army Research Office at the Edgewood Area of Aberdeen Proving Ground, Maryland. Because of the interest created by this course, a review of the Ballistic Research Laboratory's (BRL) Stepwise Multiple Regression package[1,2] was made. This review disclosed that BRL's stepwise multiple regression package was essentially developed in 1967, subject to the constraints of the 1967 hardware and therefore limited in some of the statistical tests now commonly used. This computer package is currently the workhorse in linear regression at BRL.

A recommendation by Dr. N. R. Draper[3] on linear regression statistical computer packages was requested in October 1979. Dr. Draper recommended looking at a number of commonly used statistical packages which are listed below:

- Biomedical Computer Package (BMD,[4] BMDP[5])
- Statistical Package for the Social Sciences (SPSS[6])
- International Mathematical and Statistical Library (IMSL[7])

Because of the limitations in the BRL program, this survey was undertaken to investigate a variety of regression packages including the three recommended by Dr. Draper.

---

[1] H.J. Breaux, On Stepwise Multiple Linear Regression, Ballistic Research Laboratories Report No. 1369, August 1967. (AD #658674)

[2] H.J. Breaux, L.W. Campbell, J.C. Torrey, Stepwise Multiple Regression Statistical Theory and Computer Program Description, BRL Report No. 1330, July 1966. (AD #639955)

[3] N.R. Draper, H. Smith; Applied Regression Analysis, John Wiley & Sons, Inc., 1966.

[4] W.J. Dixon; Biomedical Computer Program (BMD), University of California Press, 1973.

[5] W.J. Dixon; Biomedical Computer Program (BMDP), University of California Press, 1975.

[6] N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, D.H. Bent; Statistical Package for the Social Sciences (SPSS), McGraw-Hill, Inc., 1975.

[7] IMSL Library, Reference Manual, IMSL LIB-0007, Revised January 1979, Edition 7.

It should be understood that a vast range of computer packages are available, but that only a small subset was surveyed. This survey of linear regression packages had two purposes: (i) to familiarize users with the range of software now available and (ii) to promote an understanding of techniques used.

In Dr. Draper's text[3] Applied Regression Analysis, selected problems and their discussions were utilized as a guide in developing the list of recommended regression statistics presented below. Based upon these discussions the comparison table on several available computer packages was then developed.

II. RECOMMENDED STATISTICAL FEATURES FOR A LINEAR REGRESSION PACKAGE

Linear regression is utilized primarily to investigate relations between sets of variables and some response variable. These relations are sometimes utilized to establish predictions on a response variable. No matter how linear regression is used, this form of statistical analysis requires the calculation of associated statistics and statistical tests to evaluate the level and significance of the overall analysis. The following is a list of statistics and statistical tests which can be used to expound upon the significance of the linear regression analysis.

The first part (1-3) of the listing is simply a statement of the problem and the raw data used. The second part (4-8) is a set of statistics to compare each of the many separate regression fits to one another. The last part (9-12) evaluates the goodness of the present regression analysis for overall interpretation. This recommended list is not intended to be complete, but rather it is to be used as a guide to judge the analysis and to aid in surveying the following regression packages.

A LIST OF RECOMMENDED REGRESSION STATISTICS

1. A list of each variable and the stated regression problem.

   Regression Problem → REGRESSION ANALYSIS

   List of Variables →
   x_1 = average monthly temperature (°F)
   x_2 = number of monthly personnel
   x_3 = average monthly production (rds)
   x_4 = number of operating days in the month.
   ⋮ ⋮
   y = energy usage (dollars).

2. A listing of the original and transformed data. To check the correctness of the input data and any data transformation:

OBSERVATIONS	x_1	x_2	x_3	x_4	...	x_k	x_1'	y
1	x_{11}	x_{21}	x_{31}	x_{41}	...	x_{k1}	sin x_{11}	y_1
2	x_{12}	x_{22}	x_{32}	x_{42}	...	x_{k2}	sin x_{12}	y_2
3	x_{13}	x_{23}	x_{33}	x_{42}	...	x_{k3}	sin x_{13}	y_3
.	.	.	.	.		.	.	.
.	.	.	.	.		.	.	.
.	.	.	.	.		.	.	.
n	x_{1n}	x_{2n}	x_{3n}	x_{4n}	...	x_{kn}	sin x_{1n}	y_n

3. A list of Standard Statistics for each variable. To examine the data being analyzed.

   mean (μ_i, i=1,2, ..., k, k+1, k+2)

   $\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ $\bar{x}_4$ ... $\bar{x}_1'$ $\bar{y}$

   standard deviation (σ_i, i=1,2, ..., k, k+1, k+2)

   s_1 s_2 s_3 s_4 ... s_1' s_4

   range (maximum; minimum)

   $x_{1\,max}$ $x_{2\,max}$ $x_{3\,max}$ ... $x_{y\,max}$

   $x_{1\,min}$ $x_{2\,min}$ $x_{3\,min}$ ... $x_{y\,min}$

   correlation matrix

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1,k+2} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2,k+2} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{k+2,1} & r_{k+2,2} & r_{k+2,3} & \cdots & r_{k+2,k+2} \end{bmatrix}$$

4. The Current Regression Equation being fitted:

   $$y = f(x_i, x_j, \ldots)$$

5. The Last Variable entering the regression analysis:
   (For sequential comparisons with previous regression models)

   last variable entered = x_i .

6. A Sequential F-Test is a test to measure the significance of the entering variable into the regression equation.

7. Multiple Correlation Coefficient (R^2) is a measure of the variation being explained by current regression model.

   Percent variation explained - 42.071%

8. The standard deviation of residuals is a measure of the unexplained variation in the response variable.

9. Analysis of Variance (ANOVA) Table for regression model is a measure of the regression model relative to overall variation.

Source	df	ss	ms	Overall F
Total	N-1	SST	-	-
Regression (x_2,x_4)	2(K)	SS_R	$\frac{SS_R}{2}$	$F_{(2,N-3)}$
Residual	N-3(N-K-1)	SS_E	$\frac{SS_E}{N-3}$	

10. The estimated beta (β_i) coefficients and confidence intervals for each estimated parameter:

Var No.	$\hat{\beta}_i$ Coeff	-% Confidence Interval Upper/lower	st. error	Partial F
4	$\hat{\beta}_4$	(U_4/L_4)	st. $(\hat{\beta}_2)$	$F_{4/2}$
2	$\hat{\beta}_2$	(U_2/L_2)	st. $(\hat{\beta}_2)$	$F_{2/4}$
constant term	$\hat{\beta}_0$			

(a) The estimated coefficients of β_4 : $\hat{\beta}_4$

(b) The standard error of the estimated $\hat{\beta}_4$: $\hat{\sigma}_{\beta 4}$

(c) The confidence intervals: $\hat{\beta}_4 \pm \sigma_{\beta 4} * t(n,\alpha)$

(d) Partial F-Test: A measure of the significance of the last variable given that remaining variables are included.

11. Partial correlation of variables <u>not</u> included in current model (Regression):
A measure of the <u>remaining</u> linear correlation between the independent variables and the response variable.

12. Residual Analysis: to test the overall regression fit.

- A list of the actual observations (y_i), predictions ($\hat{y}_i$), and the difference or residuals (R_i);
- A list of $\left[\frac{y_i - \hat{y}_i}{s}\right]$ to test for normality ($N(0,\sigma_R^2)$);
- The autocorrelation function of the residuals (R_i) for independent and diagnostic testing; and
- Plot of the residuals ($y_i - \bar{y}_i$).

	y(obser)	predicted $\hat{y}$	residual $(y-\hat{y})$	N(0,1)
1	y_1	$\bar{y}_1$	R_1	$\frac{y_1-\hat{y}_1}{s}$
2	y_2	$\bar{y}_2$	R_2	$\frac{y_2-\hat{y}_2}{s}$
3	y_3	$\bar{y}_3$	R_3	$\frac{y_3-\hat{y}_3}{s}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
N	y_N	$\bar{y}_N$	R_N	$\frac{y_N-\hat{y}_N}{s}$

where $s = \sqrt{\frac{\sum_{i=1}^{N} (R_i-\bar{R})^2}{N-K-1}}$

TABLE 1. A COMPARISON TABLE OF COMPUTER PACKAGES (REGRESSION)

Characteristic	BMD[4] (1973)	BMDP[5] (1975)	IMSL[7] (1979)	SPSS[6] (1975)	SSP[8] (IBM)**	MINITAB[9] 80	DANIEL & WOOD[10] SHARE LIBRARY	ROSEPACK*,[11,12]	BRL[1,2] (H.J.BREAUX, 1968)
1) Simple Linear Regression	x	x	x	x	x	x	x	x	x
2) Multiple Linear Regression	x	x	x(RLMUL)	x	x	x	x	x	x
3) Polynomial Regression	x	x	x	L	x	x	x		
4) Non Linear Regression	x	x	x	L			x		x
5) Weighted Regression		x	x	L	x	x	x		
6) Iterative Reweighted Regression								x	
7) Other types of Regression	(a)(b)	(b)	(d)	(b) (c)(d)				x	
8) Stepwise Multiple Regression (Available Statistics)	x	x	x(RLSEP)	x	x	x	x		x
a) Standard stats for each of the variables (mean, variance, standard deviation, max, min)	x	x	x	x	x	x	L	x	
b) Skewness & Kurtosis		x		x					
c) Zero intercept option	x	x		x			x	x	x
d) Matrix data input			x	x				x	
e) Variable subset selection	x	x	x	x	L				
f) Force variables option	x	x	x	x	x		x		
g) Data transformation	x	x	(BDTRGI)	x	L	x	x	scaling & weighting	L
h) Order of variables entered	x	x	x	x	x	x	x		
i) Covariance/correlation matrix	x	x	(BECORI)	x	x	x	x		
j) ANOVA table of regression	x	x	x	x		x	x		L

k) Partial correlation	x	x		x	x	x	x		
l) Sequential F-Test	x	x	x	x	L	x	(e)		L
m) Multiple correlation coeff. (R^2)	x	x	x	x	x		x	x	x
n) Tolerance level	x	x	x(RLPRDI)	x	L				
o) Plots of residuals	x	x		x		x	x		
p) Residual analysis		x	x(RLRES)				x	x	
q) Other residual plots	x	x	x(RLRES)			x	x		
r) Lack of fit test		x	x				(e)		
s) Pure error analysis									
t) Outlier identification		(f)					x	x	
u) Missing data	x	x		x		x			
v) Other print options			x	x	L			x	L
w) Max number of variables	80	150	999	100	35	U/K	105		
x) Max number of observations	9999	9999	9999	9999	9999	core	1000	core	core
y) Programmed	FORTRAN	FORTRAN	FORTRAN	FORTRAN	FORTRAN	FORTRAN	FORTRAN	FORTRAN	FORTRAN

* - Robust Statistics Estimation Package (ROSEPACK) MIT, IMSL
** - Stepwise regression (STEPR) main program (DATA, STOUT, CORRE, MSTR, LOC, and STPRG)
x - capability exists
L - limited capability exists
(a) - periodic and harmonic regression
(b) - regression on principal components
(c) - regression with dummy variables
(d) - curvilinear orthogonal regression (RLFOR)
(e) - C Mallows variable is used at each step
(f) - Mickey, M.R., O.J. Dunn, V. Clark (1967), "Note on the Use of Stepwise Regression in Detecting Outliers." Computer & Biomedical Research 1, 105-111; and BMD Communications #1 (1974), Los Angeles, U of California.

[8] System/360 Scientific Subroutine Package (360A-CM-03X) Version III, IBM GH20-0166-S, Sixth Edition (reprinted March 1970), 1968.

[9] T.A. Ryan, Jr., B.L. Joiner, B.F. Ryan; "MINITAB, a Statistical Computing System for Students and Researchers," The Pennsylvania State University, December, 1978.

[10] C. Daniel, F.S. Wood; Fitting Equations to Data, John Wiley & Sons, Inc., 1971.

[11] D. Coleman, P. Holland, N. Kaden, V. Klema, S.C. Peters; "A System for Subroutines for Iteratively Reweighted Least Squares Computations," Massachusetts Institute of Technology (MIT), December 1977.

[12] V.C. Klema; "Robust Software for Robust Statistics," Massachusetts Institute of Technology (MIT), 1977.

III. LINEAR REGRESSION PACKAGES

Table 1 was designed with the specific purpose of summarizing the various regression subroutines that each statistical package has to offer. (See Appendix also.) However, one should be aware that in some of these program packages, there are options that allow one to obtain additional characteristics directly or indirectly. Table 1 lists the various packages and their primary characteristics.

In summary, most statistical packages are acceptable in terms of performing linear regression. In fact, with the options plus other subroutines the level and flexibility of the analysis exceeds the requirements of most users.

ACKNOWLEDGMENT

The author would like to thank Jock O. Grynovicki and James A. Matts who reviewed the draft, and especially Elizabeth A. Laurie for typing the manuscript.

REFERENCES

1. H.J. Breaux, On Stepwise Multiple Linear Regression, Ballistic Research Laboratories Report No. 1369, August 1967.

2. H.J. Breaux, L.W. Campbell, J.C. Torrey, Stepwise Multiple Regression Statistical Theory and Computer Program Description, BRL Report No. 1330, July 1966.

3. N.R. Draper, H. Smith; Applied Regression Analysis, John Wiley & Sons, Inc., 1966.

4. W.J. Dixon; Biomedical Computer Program (BMD), University of California Press, 1973.

5. W.J. Dixon; Biomedical Computer Program (BMDP), University of California Press, 1975.

6. N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, D.H. Bent; Statistical Package for the Social Sciences (SPSS), McGraw-Hill, Inc., 1975.

7. IMSL Library, Reference Manual, IMSL LIB-0007, Revised January 1979, Edition 7.

8. System/360 Scientific Subroutine Package (360A-CM-03X) Version III, IBM GH20-0166-S, Sixth Edition (reprinted March 1970), 1968.

9. T.A. Ryan, Jr., B.L. Joiner, B.F. Ryan; "MINITAB, a Statistical Computing System for Students and Researchers," The Pennsylvania State University, December, 1978.

10. C. Daniel, F.S. Wood; Fitting Equations to Data, John Wiley & Sons, Inc., 1971.

11. D. Coleman, P. Holland, N. Kaden, V. Klema, S.C. Peters; "A System for Subroutines for Iteratively Reweighted Least Squares Computations," Massachusetts Institute of Technology (MIT), December 1977.

12. V.C. Klema; "Robust Software for Robust Statistics," Massachusetts Institute of Technology (MIT), 1977.

APPENDIX. COMPUTER SOFTWARE PACKAGES

(Stepwise Multiple Linear Regression Programs)

a. Biomedical Computer Program (BMD/BMDP).

b. International Mathematical & Statistical Libraries (IMSL).

c. Statistical Package for the Social Science (SPSS).

d. IBM Scientific Subroutine Package (SSP).

e. SHARE Libraries (Daniel & Wood).

f. MINITAB 80.

g. Robust Statistical Estimation Package (ROSEPACK).

h. BRL Stepwise Multiple Linear Regression.

a. BMD (Biomedical Computer Program)

The Stepwise Regression subroutine (BMD02R) computes a sequence of multiple linear regression equations in a stepwise manner. At each step, one variable is added to the regression equation. The variable added is the one which makes the greatest reduction in the error sum of squares. In addition, variables can be forced into the regression equation. Non-forced variables are automatically removed when their F-values become too low. Regression equations with or without the zero intercept may be selected. Plots of residuals are available in this package.

BMDP (Biomedical Computer Program)

Program BMDP2R computes multiple linear regression in a stepwise manner, entering the variable that best helps to predict y into the regression equation at each step. This continues until the prediction of y does not improve notably. Whenever the correlation matrix of the predictors is singular or nearly singular the BMDP programs perform such inversion in a stepwise manner. A predictor variable is not included in the regression equation if its squared multiple correlation with the previously selected variable exceeds a certain value. Partial correlation can be computed in BMDP6R; the correlation between each pair of dependent variables is then computed after taking out the linear effects of the set of independent variables. Scatter plots of observed and predicted (expected) values of the dependent variable versus the independent variable, and plots of residuals versus other variable are available in the regression programs (R-series, Regression-Series).

TABLE A1

	BMD	BMDP
Programming Language	FORTRAN	FORTRAN
Approximate size	53 subroutines 24 K (102 K) K = 1024	26 subroutines 24 K --
No. of installations using package	Many installations	Many installations
Statistical Level	(Developed by the Dept. of Biomathematics UCLA) excellent	(Developed by the Dept. of Biomathematics) excellent
Computational Level (computer)	Health Science Computing Facility	Four Stepping Algorithms (Double Precision)
Documentation	BMD (U of CA) User's Manual	BMDP (U of CA) User's Manual
Date Developed	Jan 1973 ($8.25)	Jan 1975
Cost of package	--	$1,000.00 per year.

b. IMSL (International Mathematical and Statistical Libraries, Inc.)*

An extensive collection of mathematical and statistical subroutines written in FORTRAN. The subroutines in the regression section were designed to be useful in developing versatile application programs in the following general areas: (1) simple linear regression, (2) multiple linear regression, (3) stepwise linear regression, and (4) curvilinear regression. These 27 subroutines, integrate with other mathematical and statistical routines or functions allowing for a range from the most simple to the complex regression analysis. The system of subroutines make for a flexible system in regression analysis. IMSL is a system which aids the user in making his own programs. At each step the critical F values in subroutine RLSTP, for entering and deleting variables, change to reflect the changing error degrees of freedom. The Jordan method of

Two versions: "in-core" version, is designed to minimize usage of central processing unit time: "out-of-core" version, is designed to mimimize core storage requirements. Each of the routines calculates utilizing single and double precision.

reduction on the matrix (data) is performed. The regression package is functionally divided into two groups: (1) Linear models (RL), and (2) Special nonlinear models (RS). Subroutine RLSEP contains options for; (1) lack of fit and (2) partial F-Test (both the overall F-Test and partial F-Test for each term in the model is also performed). Routine RLSTP is the stepwise (forward) algorithm with results available after each step. The library is available in seven computer versions.

TABLE A2

IMSL

Programming Language	FORTRAN
Approximate Size	27 Subroutines
No. of Installation	Many installations
Statistical Level	Excellent
Computational Level	Subroutine RLSEP is an expanded and easy-to-use version of IMSL routine RLSTP (Double Precision)
Documentation	User's Manual
Date Developed	1977, revised January 1979
Cost	$1,220 (1 May 77) one year non-university universities, $988.00

c. SPSS (Statistical Package For the Social Sciences)

Subprogram Regression uses a forward-selection stepwise technique. Regression also allows the user to perform a regression procedure midway between two extremes by allowing the program to choose the order of introduction of the variables from a certain set, then force certain other variables into the calculation, then proceed stepwise for a period of time. There are 15 options available with subprogram Regression; (including the option for missing data; pairwise detection of missing data; ...; matrix input; output of means and standard deviation). There are seven statistics available with subprogram Regression (correlation matrix, mean, standard deviations, number of valid cases; forced printing

of the correlation matrix and removing of bad elements; ...;). Regression techniques included are: (1) curvilinear and nonadditive models, (2) regression with dummy variables, including analysis of variance and covariance models, and (3) path analysis. Assumption for nonlinear relationships (data transformation), examining polynomial trends, interaction terms etc. are included in this package. The SPSS package comes in four versions: (1) IBM OS/370, (2) CDC 6000 and CYBER 70, (3) UNIVAC 1100 series, and (4) XEROX version.

TABLE A3

	SPSS*	SCSS (controversial versions)
Programming Language	FORTRAN	FORTRAN
Approximate Size	Workspace 70,000 bytes - space allocation 80,000 bytes	UK
No. of Installations	Many	UK
Statistical Level	Excellent (One major subroutine REGRESSION)	
Computational Level	Good	Good (Double Precision)
Date Developed	1970, 1975	Fall 1979
Cost	$2,000.00†	-

*Allows for flexibility in the analysis.

†Total Package

d. SSP (IBM System/360 Scientific Subroutine Package)*

The Scientific Subroutine Package (SSP) is a set of basic computational, statistical and mathematical FORTRAN subroutines, intended to help the user develop his own packages necessary to solve problems. The package has some 250 subroutines and a number of these are in the area of regression analysis. The SSP system is best utilized with multiple

*C.A. Bennet and N.L. Franklin, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley & Sons, 1954.

use of the subroutines, such as subroutines CORRE and subroutines STPRG. There exist three modes of storage for matrix; general, symmetric, and diagonal. SSP has 15 main programs with input/output, control (parameter) cards, and sample data. Three of these main programs are regression; Regression (REGRE), polynomial regression (POLRG) and Stepwise regression (STEPR). Double-precision versions of the three subroutines are available. The Doolittle method is used in the stepwise regression subroutine. The output of the stepwise multiple regression includes: (1) for all data-means, standard deviations, and correlation coefficient matrix; (2) for each step in the multiple regression: sum of squares reduced, proportion reduced, cumulative sum of squares reduced, cumulative proportion reduced, multiple correlation coefficient (adjusted and unadjusted), F-Test for analysis of variance, standard error of estimate (adjusted and unadjusted), regression coefficients, standard errors of regression coefficients, and computer t-values; and (3) tables of residuals.

TABLE A4

SSP (Scientific Subroutine Package)

Programming Language	FORTRAN IV
Approximate Size	Over 250 FORTRAN subroutines (sample programs 32K byte (8K word))
No. of Installations	Over 300
Statistical Level	Average (standard)
Computational Level	Double Precision
Documentation	IBM publication, 1970
Date Developed	March 1970 with updated versions
Cost	Standard with IBM 360 systems (No Cost).

e. Share Lib. (Daniel & Wood):*

The linear least-square program includes options for weighting, detection of outliners and the standard analysis of variance table. The fitted equation is printed with variable names, coefficients (B(I)), t-values, minimum, maximum and range of each of the independent variables. All standard statistics are listed such as the residual root mean square, residual mean square, residual sum of squares, total sum of squares and multiple correlation coefficient squared. Also, residual values are listed with observed and predicted values plus cumulative distribution plots of residuals as standard output. The Mallows' C_p statistic is presented as one method for comparing the fitted equations. The User's Manual is available in the text (p.278) written by C. Daniel and F.S. Wood, where the restrictions** of the computer program are presented. Some twenty (20) data transformations are available as part of this linear regression program. On page 310-311 an example to measure the precision of this regression program with other commonly used least square programs is made.

TABLE A5

SHARE LIBRARY
Daniel & Wood

Programming Language	FORTRAN IV
Approximate Size	One major subroutine
No. of Installations	UK
Statistical Level	Excellent
Computational Level	Very Good
Documentation	User's Manual plus textbook (Fitting Equations to Data)
Cost	Under $100.00

Available through SHARE Library, Triangle Universities Computation Center, P.O. Box 12175, Research Triangle Park, NC 27709 (Number 360D-13.6.008).

**These restrictions may be altered by changing the dimension statements of the computer program.*

f. MINITAB 80 (Pennsylvania State University)

On-line "Help" facility, flexible transformations, interactive and batch modes, flexible plotting are standard. The MINITAB package is interactive (time sharing) as well as batch. The literature states that in a few hours, without help, a typical new user should be able to start using MINITAB. The MINITAB package contains standard correlation, regression, and Analysis of Variance. The stepwise regression and general analysis of variance are new capabilities expected in the near future (advanced version).

TABLE A6

MINITAB 80

Programming Language	FORTRAN IV
Approximate Size	Easy to install, "no difficulties." One day. (Large 80,000 words overlayed 12,000 words (48K) 20,000 lines of FORTRAN
No. of Installations	Over 300 installations
Statistical Level	Good (advanced version)
Computational Level	20,000 lines of FORTRAN; 5,000 lines of comments. (double precision)
Documentation	Well documented (Student Handbook, Reference Manual and Implementation Guide)
Date Developed	Currently being developed. PA State U, Dept. of Statistics
Cost	$1000.00 per year (new)

g. ROSEPACK (RObust Statistics Estimation Package, 1.0/2.0)

ROSEPACK is a system of portable FORTRAN subroutines to perform iteratively reweighted least squares (IRLS) robust linear regression. ROSEPACK contains 47 subroutines, a combination of numerical and statistical methods employed to optimize problems in the sense of functions of scaled residuals. Seven weighting functions are utilized in the reweighting analysis. The robust regression is aimed at analyzing and improving the behavior of least square estimation when the disturbances

are not well behaved. One goal of robust regression is to avoid undue influence on the fit if there are slight changes to all of the data or large changes to a few of the data points. The stem and leaf technique, gradient method, and orthogonal factorization are some of the methods employed in the robust regression package. Work on ROSEPACK was started in May 1975 at the Computer Research Center of the National Bureau of Economic Research and was later tested at Hampshire College, Bell Labs, and other universities (National Science Foundation Grants #DCR 75-08802, MCS 76-11989, MCS 77-12514).

The residual scaling function used in ROSEPACK is the median absolute deviation (the inclusion of other residual scaling functions is possible). The weighting functions are: (a) Huber, (b) Andrews (sine), (c) bi weight (bisquare), (d) Cauchy, (e) Welsch, (f) Talwar (zero-one), (g) Fair, (d) Logistic, and (i) user defined. The software, on tape, for the iteratively reweighted least square is available from IMSL (GNB Building, 7500 Bellaire Bldv., Houston, Texas 77036).

TABLE A7

	ROSEPACK* 1.0	ROSEPACK 2.0
Programming Language	FORTRAN	FORTRAN
Approximate Size	47 subroutines	modular, mathematical subroutines (63)
No. of Installations Using Program		Unknown
Statistical Level	Data/Numerical Analysis	
Computational Level (Computer)	(Double Precision)	
Documentation	Limited** (on-line documentation)	Limited** on-line documentation 17,500 lines of code
Developed	May 1975	March 1979
Cost	Charge for each tape $100.00	IMSL $100.00

ROSEPACK 1.0 and ROSEPACK 2.0 are available.

**ROSEPACK Staff Manager, MIT Center for Computational Research in Economic and Management Science, 575 Technology Square, Cambridge, MA 02139.*

h. HJBSLR (Harold J. Breaux; Stepwise Multiple Linear Regression).

This multiple regression is patterened after M.A. Efroymson Gauss-Jordan algorithm (Mathematical Methods for Digital Computers, John Wiley & Sons, Inc., 1960). The program utilizes a number of techniques for easing the computation. It reads and translates a formula that represents the linear model, reads the data, does the regression analysis, prints the formula that contains those terms that were finally included in the regression model. It prints the coefficients ($\hat{\beta}_i$), prints residuals if desired, and transforms the data. Confidence intervals on each regression coefficient ($\hat{\beta}_i$) are computed by the regression package.

TABLE A8

HJBSLR

Programming Language	FORTRAN IV
Approximate Size	Subroutine (regression)
No. of Installations	BRL, APG, MD
Statistical Level	Good
Computational Level	Double Precision
Documentation	BRL Reports #1330, #1369
Date Developed	1965 (updated 1966, 1967)
Cost	No cost.

DISTRIBUTION LIST

No. of Copies	Organization
12	Commander Defense Technical Info Center ATTN: DDC-DDA Cameron Station Alexandria, VA 22314
1	Commander US Army Materiel Development and Readiness Command ATTN: DRCDMD-ST 5001 Eisenhower Avenue Alexandria, VA 22333
2	Commander US Army Armament Research and Development Command ATTN: DRDAR-TSS (2 cys) Dover, NJ 07801
1	Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-LEP-L, Tech Lib Rock Island, IL 61299
1	Director US Army ARRADCOM Benet Weapons Laboratory ATTN: DRDAR-LCB-TL Watervliet, NY 12189
1	Commander US Army Aviation Research and Development Command ATTN: DRSAV-E P.O. Box 209 St. Louis, MO 61366
1	Director US Army Air Mobility Research and Development Laboratory Ames Research Center Moffett Field, CA 94035
1	Commander US Army Communications Rsch and Development Command ATTN: DRDCO-PPA-SA Fort Monmouth, NJ 07703
1	Commander US Army Electronics Research and Development Command Technical Support Activity ATTN: DELSD-L Fort Monmouth, NJ 07703
2	Commander US Army Missile Command ATTN: DRSMI-R DRSMI-YDL Redstone Arsenal, AL 35809
1	Commander US Army Tank Automotive Rsch and Development Command ATTN: DRDTA-UL Warren, MI 48090
1	Director US Army TRADOC Systems Analysis Activity ATTN: ATAA-SL, Tech Lib White Sands Missile Range NM 88002
1	University of Wisconsin Department of Mathematics ATTN: Dr. N. R. Draper Madison, WI 53706
	Aberdeen Proving Ground Dir, USAMSAA ATTN: DRXSY-D DRXSY-MP, H. Cohen Cdr, USATECOM ATTN: DRSTE-TO-F Dir, USA CSL Bldg E3516, EA ATTN: DRDAR-CLB-PA

USER EVALUATION OF REPORT

Please take a few minutes to answer the questions below; tear out this sheet, fold as indicated, staple or tape closed, and place in the mail. Your comments will provide us with information for improving future reports.

1. BRL Report Number _____

2. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)

_____

_____

_____

3. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.) _____

_____

_____

4. Has the information in this report led to any quantitative savings as far as man-hours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.

_____

_____

5. General Comments (Indicate what you think should be changed to make this report and future reports of this type more responsive to your needs, more usable, improve readability, etc.) _____

_____

_____

_____

6. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.

Name: _____

Telephone Number: _____

Organization Address: _____

_____

_____